

WHAT IS CLAIMED IS:

1. A system for encoding an audio signal, comprising:
  - 5 analyzing sequential segments of at least one digital audio signal to determine segment type as one of speech type segments, non-speech type segments, and unknown type segments;
  - encoding each speech segment as one or more signal frames using a speech segment-specific encoder;
  - 10 encoding each non-speech frame as one or more signal frames using a non-speech segment-specific encoder;
  - buffering each sequential unknown type segment in a segment buffer until analysis of a subsequent segment identifies the subsequent segment type as any of a speech segment and a silence segment; and
  - 15 encoding the buffered segments and the subsequent segment as one or more signal frames using the segment-specific encoder corresponding to the type of the subsequent segment.
2. The system of claim 1 wherein the non-speech type segments
  - 20 include silence segments and noise segments.
3. The system of claim 1 further comprising transmitting the encoded buffered segments as a burst transmission at a rate higher than a current sampling rate of the audio signal.
  - 25
4. The system of claim 1 further comprising flushing the segment buffer following each time the buffered segments and the subsequent segment are encoded.
- 30 5. The system of claim 1 wherein the sequential unknown type segments in the segment buffer are encoded using a different frame size than a

frame size used for encoding speech type segments and non-speech type segments.

5        6.        The system of claim 5 wherein the sequential unknown type segments in the segment buffer are all encoded in a single frame.

10        7.        The system of claim 5 wherein the sequential frames present in the buffer are all encoded in two frames, wherein a first frame is encoded as a speech type frame, and a second frame is encoded as a non-speech type frame.

8.        The system of claim 1 further comprising searching the sequential unknown type segments in the segment buffer to identify an actual onset point of speech corresponding to speech identified in the current segment.

15        9.        The system of claim 8 wherein the sequential frames present in the buffer are all encoded in two groups of frames, wherein a first group comprising all buffered segments preceding a segment in which the actual onset point was identified are encoded as non-speech segments, and a second group comprising the segment in which the actual onset point was identified and all subsequent  
20 buffered segments are encoded as speech segments.

10.       The system of claim 3 further comprising a decoder for receiving the burst transmission, said decoder operating at a fixed frame rate.

25        11.       The system of claim 10 wherein the decoder uses extra samples contained in the burst transmission to populate a jitter buffer.

12.       The system of claim 3 further comprising a decoder for receiving the burst transmission, said decoder using an adaptive playout scheme.

30

13. The system of claim 12 wherein the decoder uses extra samples contained in the burst transmission to populate a jitter buffer.

14. The system of claim 12 wherein the decoder compresses at least  
5 some of the received data to reduce average signal delay.

15. A system for encoding speech onset in a signal, comprising:  
continuously analyzing and encoding sequential frames of at least one  
digital audio signal while analysis of the sequential frames indicates that the  
10 sequential frames is of a frame type including any of a speech type signal frame  
and a non-speech type signal frame;

continuously analyzing and buffering sequential frames of the at least one  
digital audio signal while analysis of each sequential frame is unable to  
determine whether each sequential frame is of a frame type including any of the  
15 speech type signal frame and the non-speech type signal frame;

automatically identifying at least one of the buffered sequential frames as  
having the same type as a current sequential frame when analysis of the current  
sequential frame indicates that it is of a frame type including any of the speech  
type signal frame and the non-speech type signal frame; and  
20 encoding the buffered sequential frames.

16. The system of claim 1 further comprising temporally compressing at  
least one of the buffered sequential frames prior to encoding those frames.

25 17. The system of claim 16 further comprising searching the buffered  
sequential frames prior to temporally compressing those frames for identifying a  
speech onset point within one of the buffered sequential frames when the current  
sequential frame is a speech type signal frame.

18. The system of claim 17 wherein buffered sequential frames preceding the buffered sequential frame having the speech onset point are discarded prior to temporally compressing the buffered sequential frames.

5 19. The system of claim 18 wherein initial samples in the frame having the speech onset point which precede the speech onset point are discarded prior to temporally compressing the buffered sequential frames.

20. The system of claim 19, wherein a frame boundary of the buffered sequential frame having the speech onset point is reset to coincide with the identified speech onset point.

21. The system of claim 15 wherein the at least one digital audio signal comprises a digital communications signal.

15 22. The system of claim 15 further comprising flushing the buffer following encoding of the buffered sequential frames.

23. The system of claim 15 wherein encoding any of the sequential frames and the buffered sequential frames comprises encoding those frames using a frame type-specific encoder corresponding to the type of each frame.

24. A computer-implemented process for encoding at least one frame of a digital audio signal, comprising:

25 encoding a current frame of the audio signal when it is determined that the current frame of the audio signal includes any of speech and non-speech;

buffering the current frame of the audio signal in a frame buffer when it can not be determined whether the current frame of the audio signal includes any of speech and non-speech;

sequentially analyzing and buffering subsequent frames of the audio signal until analysis of the subsequent frames identifies a frame including any of speech and non-speech;

temporally compressing each buffered frame; and

5 encoding the temporally compressed frames as one or more signal frames.

25. The computer-implemented process of claim 24 further comprising searching the buffered subsequent frames in the frame buffer, prior to temporally  
10 compressing each buffered frame, for identifying a speech onset point within one of the buffered sequential frames when analysis of the subsequent frames identifies a frame including speech.

26. The computer-implemented process of claim 25 wherein buffered  
15 sequential frames preceding the buffered frame having the speech onset point are identified as silence frames.

27. The computer implemented process of claim 26 wherein at least one of the silence frames are discarded from the frame buffer prior to temporally  
20 compressing the buffered sequential frames.

28. The computer-implemented process of claim 24 wherein temporally compressing each buffered frame comprises applying a pitch preserving temporal compression to the buffered frames.

25

29. The computer-implemented process of claim 24 wherein temporally compressing each buffered frame comprises decimating at least one of the buffered frames.

30. The computer-implemented process of claim 24 wherein the at  
30 least one digital audio signal comprises a digital communications signal.

31. A method for capturing speech onset in a digital audio signal, comprising:

sequentially analyzing and encoding chronological frames of a digital audio signal when an analysis of the chronological frames identifies the presence  
5 of any of speech and non-speech in the frames of the digital audio signal;

buffering all chronological frames of the digital audio signal when the analysis of the chronological frames is unable to identify a presence of any of speech and non-speech in the frames of the digital audio signal;

identifying at least one of the buffered chronological frames as having a  
10 same content type as a current chronological frame of the digital audio signal when the analysis the current chronological frame identifies the presence of any of speech and non-speech in the digital signal following the buffering of any chronological frames; and

encoding the current chronological frame and at least one of the buffered  
15 chronological frames.

32. The method of claim 31 further comprising temporally compressing at least one of the buffered frames when the analysis of the chronological frames prior to encoding the current chronological frame and at least one of the buffered  
20 chronological frames.

33. The method of 32 further comprising searching the buffered chronological frames in the frame buffer, prior to temporally compressing at least one of the buffered chronological frames, for identifying a speech onset point  
25 within one of the buffered chronological frames, and wherein said search is initialized using speech identified in the current chronological frame.

34. The method of claim 33 wherein buffered chronological frames preceding the buffered chronological frame having the speech onset point are  
30 identified as non-speech frames.

35. The method of claim 33 wherein samples of the at least one digital audio signal within the buffered chronological frame having the speech onset point are identified as non-speech samples.

5           36. The method of claim 31 wherein the at least one digital audio signal comprises a digital communications signal in a real-time communications device.